B1

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/00, 21/04, C07K 2/00, 14/00, C12N 5/10, 5/16, 7/01, 15/00, 15/10, 15/11, 15/12, 15/63, C12P 1/00, 21/00, 21/02 — A1

(11) International Publication Number: WO 95/18140

(43) International Publicati n Date: 6 July 1995 (06.07.95)

(21) International Application Number: PCT/US94/14554

(22) International Filing Date: 28 December 1994 (28.12.94)

(30) Priority Data:
108205 28 December 1993 (28.12.93) IL

(71) Applicant *(for all designated States except US)*: YEDA RESEARCH & DEVELOPMENT CO., LTD. [IL/IL]; Weizmann Institute of Science, P.O. Box 95, 76100 Rehovot (IL).

(71) Applicant *(for VN only)*: RYCUS, Avigail [US/IL]; 16 Kipnis Street, 76305 Rehovot (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BEN-ARIE, Nissim [IL/IL]; 105 Eshkol Street, 76804 Bazkeret Batya (IL). LANCET, Doron [IL/IL]; 4A Melchett Street, 65215 Tel Aviv (IL).

(74) Agents: COHEN, Herbert et al.; Wigman, Cohen, Leitner & Myers, P.C., Suite 200, Crystal Square 3, 1735 Jefferson Davis Highway, Arlington, VA 22202 (US).

(81) Designated States: CA, JP, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: OLFACTORY GENES AND RECEPTORS

(57) Abstract

Olfactory genes appear as gene clusters. In the human geriome such clusters appear on chromosomes 11 and 17. Such gene clusters are disclosed herein for the first time. A variety of DNA sequences encoding olfactory receptors or parts thereof have been found and are disclosed herein.

10/024,399 # 11

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

OLFACTORY GENES AND RECEPTORS

FIELD OF THE INVENTION

The present invention is generally in the field of human genetics and concerns genes of the human olfactory system.

PRIOR ART

The following is a list of prior art referred to in the present specification:

1. Buck, L. and Axel, R., 1991, *Cell* 65:175–187.
2. PCT Application WO 92/17585.
3. Levy, N.S., *et al.*, 1991, *J. Steroid Biochem. Molec. Biol.*, 39:633–637.
4. Buck, L.B., 1992, *Current Opinion Genet. Devel.*, 2:467–473.
5. Parmentier, M., *et al.*, 1992, *Nature*, 355:453–455.
6. Nef, P., *et al.*, 1992, *Proc. Natl. Acad. Sci. USA*, 89:8948–8952.
7. Selbie, L.A., *et al.*, 1992, *Mol. Brain Res.*, 13:159–163.
8. Raming, K., *et al.*, 1993, *Nature*, 361:353–356.

9. Ressler, K.J., *et al.*, 1993, *Cell*, 73: 597–609.

10. Ngai, J., *et al.*, 1993, *Cell*, 72:657–666.

11. Schurmans, S., *et al.*, 1993, *Cytigenet Cell Genet*, 63:200–204.

12. Nizetic, D., *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88:3233–3237.

13. Kwiatkowski, D.J., *et al.*, 1990, *Am. J. Hum. Genet*, 46:559–567.

14. Pinkel, D., *et al.*, 1986, *Proc. Natl. Acad. Sci. USA*, 83:2934–2939.

15. Fidlerová, H., *et al.*, 1993, *Submitted.*

16. Ledbetter, D.H., *et al.*, 1989, *Proc. Natl. Acad. Sci. USA*, 86:5136–5140.

17. Rackwitz, H.R., *et al.*, 1985, *Gene*, 40:259–266.

The acknowledgement herein of the above publications is given to allow an appreciation of the prior art but should not, however, be construed as an indication that this art is in any way relevant to the patentability of the invention as defined in the appended claims.

The above publications will be acknowledged herein by indicating their number from the above list.

BACKGROUND OF THE INVENTION

The genetic basis for the olfactory system has been subject to extensive research. An understanding of the genetic basis of olfaction is important, for example, for the fragrance industry, to enable a more systematic and accurate way of fragrance design. To date, research in this field has been centered primarily in non–human mammals, although some information on the genetic basis of the olfactory system in humans has also been found.

On the basis of growing body of evidence, it became clear that the olfactory receptors (ORs) are encoded by a superfamily consisting of a

plurality of genes[1-9]. Common characteristics of all receptors identified to date is that they belong to the superfamily of G-protein coupled receptors and have seven putative trans-membrane helices.

Genomic blot analyses have shown that the rat genome may have several hundred OR genes. It is anticipated that large OR gene repertoires exist also in other mammals. In addition to the study of OR genes in rats[1,3,8], members of the OR gene family have also been cloned from fish[10], mouse[6,9], dog[5] and human[5,7,11].

It is an object of the present invention to provide novel DNA molecules encoding human olfactory receptors.

It is more specifically an object of the invention to provide a human olfactory subgenome, i.e. a cluster of OR genes.

It is another specific object of the present invention to provide novel proteins or polypeptides encoded by the DNA sequences of the invention.

The remaining objects of the invention will be realized in the following description and claims.

GENERAL DESCRIPTION OF THE INVENTION

In accordance with the present invention it has been found that contrary to the prior art prevailing belief that known OR genes are randomly distributed throughout the genome, human OR genes form continuous clusters in the genome (hereinafter *"OR gene clusters"*). One such OR gene cluster was found on human chromosome 17. This OR gene cluster was found to be about 0.35 Mb long with an average inter-genic separation between the genes of about 15 kb.

Furthermore, in accordance with the present invention novel DNA coding sequences from OR genes belonging to the OR cluster on chromosome 17 have been identified and cloned. These sequences are

depicted in Fig. 4 which shows 16 sequences of which those designated 17–2 and 17–4 are known and the rest are novel.

The present invention thus provides, by a first of its aspects, a DNA molecule selected from the group consisting of:

(a) an isolated cluster of olfactory receptor genes;

(b) an isolated non–coding fragment of the DNA sequence of (a);

(c) a DNA sequence included in (a) or (b) and included within a sequence coding for an olfactory receptor with the proviso that the sequences designated 17–2 and 17–4 in Fig. 4 are excluded;

(d) a DNA sequence different than the DNA sequence of (c) and encoding the same protein or polypeptide;

(e) a DNA sequence of (c) or (d) in which one or more coding triplets from the open reading frame have been added, deleted or replaced, the protein or polypeptide encoded by said sequence having similar properties to that encoded by the DNA sequence of (c) or (d); and

(f) a DNA sequence capable of hybridizing under conditions of stringency to the DNA sequence of (a), (b), (c), (d) or (e).

A specific olfactory receptor gene cluster in accordance with the invention is that located on human chromosome 17 or that located on human chromosome 11.

By a further aspect of the invention there is provided a DNA molecule having a coding sequence being a member selected from the group consisting of:

(a) a DNA sequence depicted in Fig. 4 with the proviso that the sequences designated 17–2 and 17–4 be excluded;

(b) a DNA sequence encoding a functional polypeptide being a fragment of the polypeptide encoded by the DNA sequence of (a);

(c) a DNA sequence comprising the DNA sequence of (a) or (b);

(d) a DNA sequence of (a), (b) or (c) wherein one or more triplets in the open reading frame have been added, replaced or deleted, the protein or polypeptide encoded by the sequence having essentially the same properties as the protein encoded by the respective sequence of (a), (b) or (c);

(e) a DNA sequence of (a), (b), (c) and (d) comprising also non-coding stretches of nucleic acid residues; and

(f) a DNA sequence capable of hybridization under conditions of stringency with the DNA sequence of (a), (b), (c) or (d).

The above DNA molecules can be used in transforming a host cell to express the protein or polypeptide encoded by said DNA molecule. For this purpose, the DNA molecule is incorporated into an expression vector which is introduced into suitable host cells, which may be prokaryotic or eukaryotic. Eukaryotic host cells are generally preferred since the post-translational modification in prokaryotic and eukaryotic host cells is different, and thus expression in a eukaryotic host cell will bring to an expression product which more closely resembles that encoded by the olfactory genomic DNA in humans. A particular example of a vector and a host cell is the Baculovirus vector and an insect host cell, respectively.

The present invention thus provides a method of production of a protein or a polypeptide in a host cell comprising transforming a host cell by a DNA molecule selected from those defined above, growing the host cells under conditions allowing for expression of the protein or polypeptide encoded by said DNA molecule and harvesting the expressed product.

In accordance with yet another aspect of the invention, there is provided a protein or polypeptide encoded by the above coding sequence and being a member selected from the group consisting of:

(a) a protein or polypeptide having an amino acid sequence depicted in Fig. 3;

(b) a protein or polypeptide wherein one or more amino acid residues has been added, deleted or chemically modified, the protein or polypeptide having essentially the same biological properties as the respective protein or polypeptide under (a); and

(c) a protein comprising a polypeptide having a sequence of (a) or (b).

The present invention also provides, by a further of its aspects, uses of the above DNA molecules, proteins or polypeptides.

Another use of the DNA molecule is in the construction of probes for use in studying olfactory polymorphism in a population or of probes for Southern blot hybridization and restriction fragment length polymorphism for detecting interindividual differences in olfactory receptors or olfactory clusters in humans. Useful for this purpose are characterizing fragments of an OR gene cluster or DNA sequence which hybridize thereto. The integrity of gene clusters is usually preserved over generations, i.e. gene clusters are usually inherited as a single unit. This is so since the recombination between pairs of chromosomes has a much higher probability of occurring outside such clusters. In such clusters, unique non-coding sequences of a length of a few tens of nucleic acid basis can usually be found, which can be used to identify the entire cluster. By the use of such fragments or nucleic acid sequences which can hybridize thereto, the distribution of OR gene cluster in a population and hence OR polymorphism, can be assayed.

Probes to assay for OR polymorphism are very important in the fragrance industry since on the basis of knowledge of such polymorphism, fragrances for use by specific populations may be used. Based on the type of olfactory receptors, different people may be responsive to different fragrances. At times, a homogeneous population may have many common

ORs and thus fragrance preparations specific for this population may be produced. Alternatively, by the knowledge of polymorphism in a population, a fragrance preparation can be produced which will create a positive response in most individuals.

The proteins or polypeptides of the invention may be used in probing for the presence of a certain fragrance, aroma or flavor in a medium. Another use of said polypeptides or proteins is in the development of fragrances for human use. When used in probing or in the development of fragrances, said protein or polypeptide may be immobilized onto a solid support or may be embedded into a suitable membraneous system, e.g. membrane of a host cell transformed with a DNA encoding for said protein or polypeptide displaying same on its surface, a liposome wherein said protein or polypeptide is embedded, and the like. In such use, binding of a fragrance, aroma or a flavor to said protein or polypeptide will give rise to a detectable signal. The manner of use of a receptor protein so that binding of a ligand thereto will give rise to a detectable signal is generally known *per se* and the description of such manner is outside the scope of the present writing.

A further use of the above protein and polypeptide is in raising antibodies. Such antibodies, which also form an aspect of the invention, may be used to detect and label ORs *in vivo* as well as in *in vitro* preparations. For such use, the antibodies may carry a detectable marker. Another use of the antibodies is in assays for the detection of the presence of fragrances in a medium. In such use, antibodies which recognize the fragrance binding site are brought into contact with said proteins or polypeptides in the presence of a tested medium, the level of bound or free antibodies being directly proportional to the level of fragrance in the tested medium.

As will no doubt be appreciated by the artisan, the DNA molecules and proteins or polypeptides of the invention may have various other uses, all being within the scope of the present invention.

The invention will now be illustrated in the following description of specific embodiments with occasional reference to the annexed drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows fluorescence *in situ* hybridization (FISH) of cosmid 7 to metaphase spreads of human chromosomes. The arrows in the figure indicate labelling at the telomeric end of chromosome 17 on the two sister chromatids of both chromosomes.

Fig. 2 shows the mapping of the OR cluster in human 17p13.3. Thirty–six representative OR–positive cosmids from the human chromosome 17 library were assembled into 3 contigs utilizing a partial digest mapping technique. The *Hind*III (H) and *Eco*RI (E) sites are shown on the top lines. The numbered cosmids are shown as horizontal line, showing their positions in the contig. Gap lengths between the 3 contigs were estimated to be 10–20 kb by FISH to free chromatin, and are shown to scale. OR coding regions mapped to within ±4kb are shown as dark boxes, while those for which only a cosmid assignment is known are shown as open boxes. The approximate location of STS 506 (D17S126) and the PCR marker HP5 (in the 3′ non coding region of OR17–OR17–210) are also shown. None of the cosmids tested positively with the flanking chromosome 17 STS's PFN1 and 508[13]. The approximate DNA breakpoints of Miller–Dieker Syndrome patient BR8 is shown, within cosmid 7 by Southern blot analysis of DNA from a cell line containing chromosome 17 of this patient on mouse background, with radiolabelled cosmid 7 DNA;

Fig. 3 shows the deduced amino acid sequences of olfactory receptor clones from human chromosome 17 and their alignment. The deduced

protein sequences of the PCR-generated clones encoding putative human ORs were aligned using the GCG package pileup program (Genetic Computer Group, Madison, Wisconsin, U.S.A.). Rat OR I14[1] is also included for comparison and positioning of the sequences in the full length open reading frame. The numbering system used is based on the I14 sequence. Transmembrane domains were determined by an averaged hydropathy plot procedure combined with a hydrophobic moment analysis and are marked by lines with roman numerals. Dark boxes represent amino acids identity, while different shades of gray stand for two different degrees of conservative substitutions. Marked by a ● are positions of full identity among all the sequences. ORs 17-23 and 17-90 have a deletion of one nucleotide compared to all other sequences, which leads to an immediate stop codon to be in frame. Therefore, amino acid 166 in these sequences was marked as *, and the remainder of the protein is the translation of the phase deduced by homology to the other OR sequences;

Fig. 4 shows the DNA sequences of the olfactory receptors from the human chromosome 17 cluster shown in Fig. 3; and

Fig. 5 shows the expression of OR17-93 in human olfactory epithelium. PCR analysis was performed using OR17-93 specific primers on the following templates; OR17-93 and OR17-2 clones as positive controls (OR), cDNA of olfactory epithelium (OR), cDNA of respiratory epithelium (RE), cDNA from nasal inverted papilloma (PA). Arrows are pointing at the PCR products of OR17-93 and OR17-2 clones (260 and 230 bp, respectively). M=100 bp ladder markers (BRL, Gaithersburg, MD).

DESCRIPTION OF A SPECIFIC EMBODIMENT

The invention will now be described by outlining the experiments which led to the discoveries which form the basis of the present invention.

This section is intended for illustration of the invention and should not be construed to limit the invention in any way.

Materials and Methods

OR Probe preparation

Genomic DNA was extracted from male Wistar rats and from blood of a human individual. The rat probe (ORI) was obtained by PCR using oligonucleotides OR5B (TM2; CCC ATG TA(T/C) TT(G/C/T) TT(C/T) CTC (A/G/T)(G/C) (C/T) AA(C/T)(T/C)T(G/A) TC C) and OR3B (TM7; OR3B AG(A/G) C(A/T)(A/G) TAI ATG AAI GG(A/G) TTC AIC AT). The human probe, ORIII, was amplified by nested PCR, using first OR5B/OR3B and then reamplified by OR5A (GGC (C/T)TA TGA CCG IT(A/T) (T/C)(G/C)T (G/A)GC (C/T)AT ITG) and OR3B. PCR amplification was carried out in a buffer containing 50mM KCl, 10mM Tris pH 8.3, 0.01% gelatine, 0.2 mM each deoxyribonucleotide, 1μM each primer and 2.5 U of Taq DNA polymerase (Cetus, Branchburg, NJ) 100μl as follows: 30 cycles of denaturation at 94°C, annealing at 55°C and extension at 72°C, each step for 1 min. The first step of denaturation and the last step of extension were 4 min. long. Both PCR-generated probes were purified by Geneclean II (BIO 101, La Jolla, CA).

Library Screening

A 5-fold coverage Chromosome 17 cosmid library was used, prepared previously from flow-sorted chromosome 17 DNA, isolated from human cell line LCL127[12]. DNA probes were radiolabelled to a specificity activity of 10^8-10^9 cpm/μg by random hexamer priming using [α-32P] dCTP (Amersham, Buckinghamshire, UK). Hybridization to filters was carried out overnight in 50% formamide, 50 mM sodium phosphate buffer, pH 7.2, 1 mM EDTA, 10X Denhardt's, 4X SSC, 1% SDS, 10% dextran

sulphate, 50 μg/ml sonicated salmon sperm DNA, and 100 μg/ml sonicated human placental DNA at 42°C. Cosmid and genomic filters were washed in 1X SSC, 0.1% SDS at room temperature for 60 minutes and exposed to Kodak XAR 5 X-ray film at -80°C. The positive cosmids isolated were numbered sequentially from 1 to 76. The isolation of 70 OR-positive cosmids (average insert length 35 kb) in a region of 350 kb suggests a 7-fold coverage of the library, over this region, in agreement with the previously estimated coverage.

'Partial digest' mapping of the cosmids

Cosmid DNA was purified using a Qiagen column (Diagen, Chatsworth, CA) according to the manufacturer's recommendations. For the cosmid blots, 1 μg of DNA was digested by the appropriate restriction enzyme and 400 ng run on an 0.65% agarose gel. The DNA was visualized by staining with ethidium bromide and transferred onto nylon membranes (Hybond N+, Amersham, Buckinghamshire, UK), according to the manufacturer's recommendations. For partial digest mapping, 10 ug of cosmid DNA was linearised by cleavage with lambda terminase (1U/mg DNA) for 2 hours at room temperature and 2 μg partially digested with *Eco*RI or *Hind*III in a total volume of 30 μl at 37°C (0.2 U/μg DNA, removing 5 μl aliquots every 5 minutes). Partially digested cosmid DNA were hybridized in solution with radiolabelled oligonucleotides complementary to the protruding 12bp single-stranded sequences at the cleaved cos site and the samples fractionated on a 0.5% agarose gel. The annealing of labelled oligonucleotides specific for the 'left' or 'right' termini of the cos sites allowed the autoradiographic visualization of restriction fragments of increasing length extending from each side of the cosmid. A restriction map was constructed by calculating and compiling the digitized migration distances of partial digestion restriction fragments.

Minimal Complexity Index (MCI) Analysis

In order to find how many OR coding regions lie in each cosmid, PCR amplification using OR5B and OR3B as primers was performed on individual cosmids. The PCR products were digested with a frequent cutter restriction enzymes (e.g. *Hinf*I, *Hae*III, *Hha*I). After agarose gel separation and ethidium bromide staining, the lengths of the fragments were summed and divided by the length of the undigested product, as described [1,2]. The value obtained is the Minimal Complexity Index (MCI), indicating the minimal number of putative receptors encoded by the cosmid.

Sequencing of cloned PCR products

Cosmid DNA was extracted by the miniprep alkaline lysis procedure. PCR was performed separately on each, using OR5B/OR3B primers. An equimolar pool of PCR products from all cosmids, as well as products amplified from individual cosmids were ligated to pCR1000 or pCRII plasmids (Invitrogene, San Diego, CA). Recombinant clones were identified by Southern blot hybridization and PCR. Sequencing was performed using Taq DNA polymerase and dye-terminators on an ABI DNA sequencer model 373A, with universal vector primers and HOR1, an internal primer (GA/(C/T) G(A/G)(A/C) T(A/T)(C/T) (G/C)TG GCC ATG TGC).

Fluorescence *in situ* hybridization (FISH)

Hybridization and detection were carried out according to a modification of the technique described by Pinkel *et al.*[14]. For metaphase *in situ* hybridization, 1 μg of cosmid DNA was biotinylated using a Bionick kit (BRL, Gaithersburg, MD). 80 ng of labelled DNA was mixed with 4 μg Cot-1 DNA (BRL, Gaithersburg, MD), dried and resuspended in hybridization buffer (50% formamide, 2xSSC, 10% dextran sulphate, 1% Tween 20).

The sample was denatured by heating to 80°C for 3.5 minutes and incubated at 37°C for 60 minutes. Chromosome slides were baked for two hours at 65°C and then denatured in 70% formamide, 2xSSC for 3 minutes at 75°C. Slides were dehydrated by washing with 70%, 95% and absolute ethanol. Hybridization took place overnight at 37°C under sealed coverslip. Post hybridization washes were with 2xSSC. 50% formamide (3 x 5 min.) and 2xSSC (3 x 5 min.) both at 42°C. The slides were then washed in 4xSSC, 0.05% Tween 20, pH 7 (SSCT) for 3 minutes at 37°C and preincubated in SSCT containing 5% low fat dried milk (Marvel) (SSCTM) prior to the addition of FITC–conjugated avidin DCS (Vector labs) at 5 μg/ml in SSCT. The signal was amplified by addition of biotinylated anti–avidin (5 μg/ml) and a second application of FITC–avidin (DCS (5 μg/ml). Finally the slides were washed with phosphate buffered saline and dehydrated through an ethanol series.

Chromatin was released from routinely fixed nuclei according to a recently developed method[15]. For two color FISH one probe was labelled with biotin and one with digoxygenin, as above. Detection was by incubation with mouse anti–digoxygenin monoclonal antibody (Boehringer Manheim, Manheim, Germany) at 0.8 μg/ml plus avidin Texas red at 2 μg/ml, both in SSCTM. This was followed by anti–mouse IgG conjugated to FITC at 30 μg/ml. The biotinated anti–avidin was used alone and finally another round of avidin Texas red was performed. Post hybridization washes were as described. Slides were mounted in Citiflour (Citiflour Ltd., London) with propidium iodide 0.5 μg/ml for one color visualization or DAPI (diamidino–2–phenyl–indole dihydrochloride; Sigma, St. Louis, MO) 0.1 μg/ml for two colors, and photographed using a Zeiss Axiophot microscope.

Somatic cell hybrid mapping

Somatic cell hybrids BR8 and AY1 were constructed by fusion of lymphoblastoid cell lines of Miller–Dieker syndrome patients MDS–9 and MDS–12, respectively, with TK′ rodent cells followed by selection in HAT media. Hybrid clones containing a single, detected chromosome 17 were identified by RLFP analysis of polymorphic loci in distal 17p and confirmed by cytogenetic analysis.

OR tissue expression

Human tissues were surgically removed from a female patient having inverted papilloma and were immediately frozen in liquid nitrogen. Total RNA was extracted using TRI–REAGENT (NRC Inc., Cincinnati, OH) following the manufacturer's procedure. Yields were 3.3, 0.9 and 1.6 mg RNA/g tissue for papilloma, olfactory epithelium and respiratory epithelium, respectively. Total RNA (2 μg) was treated with 0.2 units reverse transcriptase (BRL, Gaithersburg, MD) for 20 mins. at 37°C and was amplified using OR17–93 specific primers (OR17–93A: ATT GCT GCA GAA CAT GCA GAA; OR17–93B: ACA TGG CAT GGA AGG TGG TC). A control experiment was conducted to verify that the PCR signals are RNA–dependent, and not due to a genomic DNA contamination. Since Taq DNA polymerase is known to have a significant reverse transcriptase activity, the omission of reverse transcription was considered less desirable. Instead, control PCR reactions were performed using primers for β–actin spanning intron 2.

Results

OR–positive cosmids from human chromosome 17

Conserved regions in the 2nd and 7th transmembrane domains of the published rat OR sequences[1,2] were used to design oligonucleotide

prime respectively). These were utilized to amplify a compl sequences by polymerase chain reaction (PCR), using nic DNA as templates (probes ORI and ORIII, respect g PCR products were radiolabelled and used to screen ared from flow-sorted human chromosomes 17, 21, 22 number of positively hybridizing clones was obtained me 17: 75 cosmids out of approximately 20,000 tested g ization signal with the human genomic probe (ORIII), (36) contained OR genes sufficiently conserved to be de I rat PCR probe as well. The clones were individua presence of OR coding sequences by Southern hybridizat mplification with the primers described above. The latter e OR genes have no introns in their coding regions[1,2,6] zation positive cosmids gave the expected ~ 700 bp long p by the primers OR5B and OR3B, and were further cha of the PCR products was then digested with frequent-cu nzymes, and subjected to a Minimal Complexity Index (MCI Materials and Methods), giving an estimate for the minimal oding regions per cosmid. The results showed an average o 35 kb cosmid genomic insert, or approximately an average o g region per 15 kb of genomic DNA. There was a correlat e intensity of hybridization signal in the library screen and th nber of ORs present on a given cosmid (not shown).

Six ing disparate OR restriction patterns (numbers 1,2,4,7,8,10) w luorescence *in situ* hybridization to metaphase spreads of hum mes (Fig. 1). All four mapped to a similar position, 17p13 p-telomere of chromosome 17. This result, together with th esence of more than one OR coding region per

cosmid provided initial indications that at least some of the identified OR genes were clustered.

Chromosomal mapping of the OR gene cluster

The approximate size of the genomic OR locus was determined by hybridizing the 5B/3A PCR products from cosmids, 1,2,4 and 7 to a Southern blot containing human genomic DNA digested with a variety of infrequently cutting restriction enzymes (*BssHII,* SacII, *Mlu*I, *Nru*I and *Not*I). All four PCR products identified a ~ 400 kb *Bss*HII restriction fragment, giving an estimate for the maximal size of the region. This indicated that the clustered OR genes do not have CpG islands at their 5′ flanking ends. *Hind*III and *Eco*RI restriction maps were constructed for 40 cosmids using a partial digest mapping protocol[17], and 35 of them arranged into three contigs with a total length of 347 kb (Fig. 2). All the other cosmids were judged by full restriction digest patterns with *Hind*III to be overlapping with the mapped subset. Two small gaps were seen, which probably arise due to the large distance between neighboring OR genes in these regions. The gap sizes between cosmids 53–45, 53–68 and 28–65 were estimated at 10–20 kb, and the contigs oriented relative to each other, by two-color FISH[15] to free chromatin using pairwise combination of cosmids from the ends of the three established contigs (data not shown). This produced clear linear signals from cosmids hybridized to chromatin fibers, enabling the relationship between adjacent or overlapping probes to be determined.

A more accurate localization and an absolute orientation of the OR cluster within the chromosomal band 17p13.3 was obtained using DNA from somatic cell hybrids containing the deleted human chromosome 17 homolog of two Miller–Dieker Syndrome (MDS) patients[16]. DNA from hybrid BR8 was positive with the OR17–93 PCR primers but negative with

PCR marker HP5 (near OR17–210) as well as with STS marker 506[16] which is found only on cosmid 73 (Fig. 2). Southern hybridization with cosmids 7, 44 and 59 further confirmed the location of the BR8 breakpoint at the centromeric end of the OR cluster, within cosmid 7. DNA from hybrid AY1 was positive with all three PCR markers, consistent with its breakpoint being telomeric to STS marker 506. Thus, the OR cluster is localized between the BR8 breakpoint and STS 506, immediately centromeric to the MDS critical region.

Mapping and sequencing of OR coding regions

In order to localize the OR genes in the cluster, and to obtain DNA sequence information for them, an equimolar mixture of coding regions amplified by PCR (with primers OR5B–OR3B) from each of the 70 cosmids, was used to generate a library of OR clones. Of 17 clones sequenced, 11 were different. Restriction patterns of OR5B–OR3B PCR products with up to six frequent-cutting restriction enzymes allowed a comparison of these clones to those seen in PCR products from the individual cosmids. This suggested the existence of 6 additional OR genes in the cluster, and provided cosmid assignments for each of the 16 OR coding regions identified. By relating this information to the cosmid physical map, it was possible to localize most of the receptors with an accuracy of ±4 kb within the cluster. Further positional information was obtained from Southern blot hybridization of the ORIII probe to *Hind*III, and *Hind*III/*Eco*Ri digests of the cosmids (Fig. 2).

The 17 OR protein sequences (between transmembrane domains 2 and 7) are shown in Fig. 3 and the corresponding DNA sequences are shown in Fig. 4. The ORs have 36–39% mutual amino acid sequence identity. Two of the sequences (OR17–2 and OR17–4) are identical to recently published human OR clones HUMHGM071[5] and HGM07E[11],

respectively. The latter gene was independently mapped to 17p13-->p12[11]. All the OR gene sequences were co-linear, except OR17-93 which had a 9 amino acid insert between positions 129 and 130, within intracellular loop 2. In addition, there is a single amino acid deletion at position 268 in four of the OR sequences. In two cases, pairs of OR genes shared >99% identity, with members of each pair mapping to the same position within the accuracy of our determination. For the first pair (OR17-2 and OR17-31), the difference was one silent mutation (at amino acid 218). For the second pair (OR17-23 and OR17-90, both pseudogenes with the same single base deletion at amino acid 166) the difference was a trinucleotide deletion in OR17-90, which affected two coded amino acids (positions 243-4). These two OR gene pairs may be alleles, occurring at respectively identical loci on cosmids derived from the parental and maternal chromosomes. Alternatively, they could be the result of a recent tandem duplication.

Tissue expression

The OR genes described here were isolated from cloned genomic DNA fragments. To examine whether any of them are actually expressed in human olfactory epithelium, samples of olfactory and non-sensory nasal tissues from a human subject were obtained. Total RNA was extracted, and examined for OR expression by a reverse transcriptase - polymerase chain reaction (RT-PCR) assay (Fig. 5). A control PCR with β-actin primers spanning an intron was used to demonstrate that signals are not significantly generated from a genomic DNA contamination. Since OR genes are rather similar to each other, it was necessary to verify that a specific OR of the presently described cluster is the one actually amplified. The verification made use of the unusual 27 nucleotide insertion in OR17-93 and this, together with the design of a specific primer pair allowed to ascertain that only one gene product was amplified.

CLAIMS:

1. A DNA molecule selected from the group consisting of:
   (a) an isolated cluster of olfactory receptor genes;
   (b) an isolated non-coding fragment of the DNA sequence of (a);
   (c) a DNA sequence included in (a) or (b) and included within a sequence coding for an olfactory receptor with the proviso that the sequences designated 17-2 and 17-4 in Fig. 4 are excluded;
   (d) a DNA sequence different than the DNA sequence of (c) and encoding the same protein or polypeptide;
   (e) a DNA sequence of (c) or (d) in which one or more coding triplets from the open reading frame have been added, deleted or replaced, the protein or polypeptide encoded by said sequence having similar properties to that encoded by the DNA sequence of (c) or (d); and
   (f) a DNA sequence capable of hybridizing under conditions of stringency to the DNA sequence of (a), (b), (c), (d) or (e).

2. A DNA molecule according to Claim 1, wherein the isolated gene cluster is derived from human chromosome 11 or 17.

3. A DNA molecule having a coding sequence being a member selected from the group consisting of:
   (a) a DNA sequence depicted in Fig. 4 with the proviso that the sequences designated 17-2 and 17-4 be excluded;
   (b) a DNA sequence encoding a functional polypeptide being a fragment of the polypeptide encoded by the DNA sequence of (a);
   (c) a DNA sequence comprising the DNA sequence of (a) or (b);
   (d) a DNA sequence of (a), (b) or (c) wherein one or more triplets in the open reading frame have been added, replaced or deleted, the protein or polypeptide encoded by the sequence having

essentially the same properties as the protein encoded by the respective sequence of (a), (b) or (c);

(e) a DNA sequence of (a), (b), (c) and (d) comprising also non-coding stretches of nucleic acid residues; and

(f) a DNA sequence capable of hybridization under conditions of stringency with the DNA sequence of (a), (b), (c) or (d).

4. A method of production of a protein or a polypeptide in a host cell comprising transforming a host cell by a DNA molecule of Claim 3, growing the host cells under conditions allowing for expression of the protein or polypeptide encoded by said DNA molecule and harvesting the expressed product.

5. A protein or polypeptide being a member selected from the group consisting of:

(a) a protein or polypeptide having an amino acid sequence depicted in Fig. 3;

(b) a protein or polypeptide wherein one or more amino acid residues has been added, deleted or chemically modified, the protein or polypeptide having essentially the same biological properties as the respective protein or polypeptide under (a); and

(c) a protein comprising a polypeptide having a sequence of (a) or (b).

6. A host cell transformed by a DNA molecule according to any one of Claims 1 or 3.

7. A probe for use in assaying fragrances, aromas or flavors in a medium, comprising a protein or polypeptide according to Claim 5.

8. A probe for use in testing of the presence of a gene cluster among individuals in a population, comprising a DNA molecule as defined in Claim 1(b).

9. An antibody directed against a protein or polypeptide according to Claim 5.

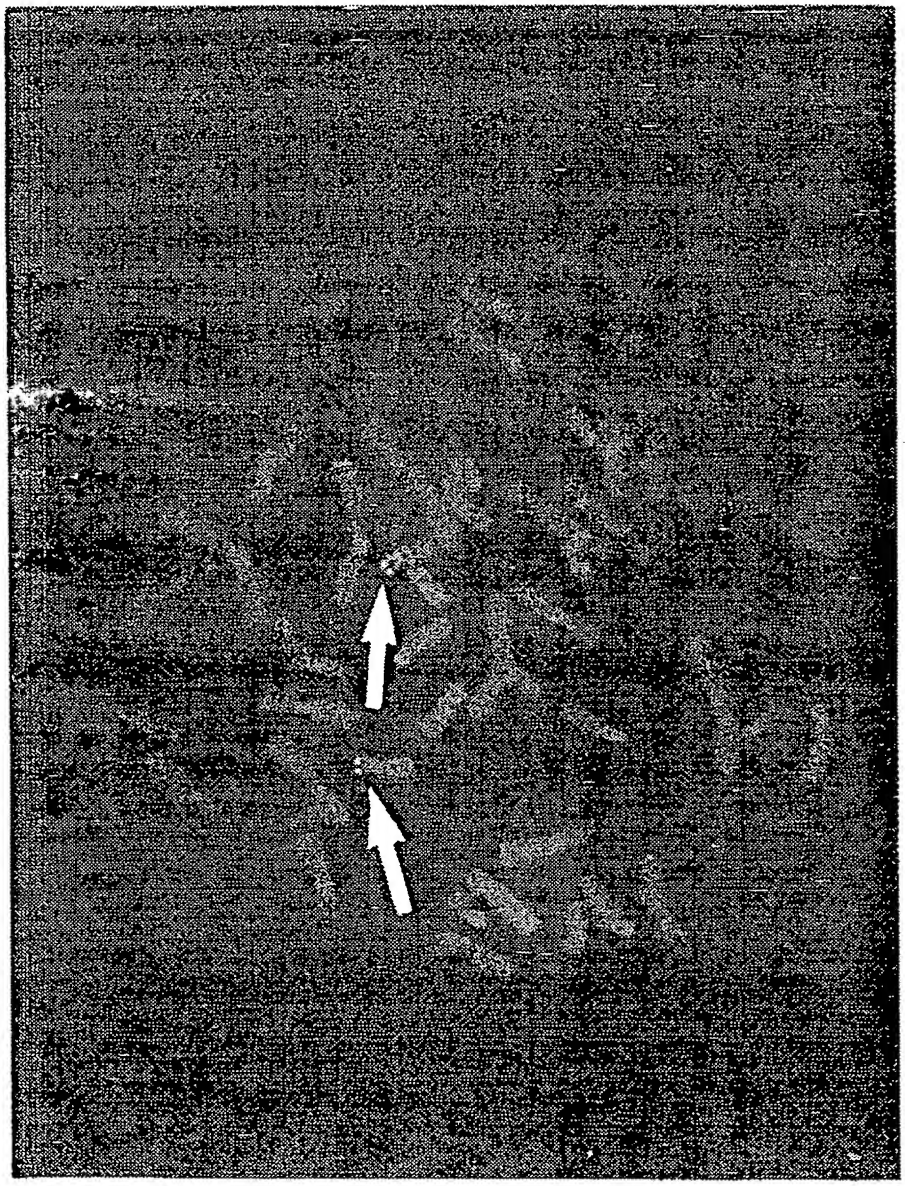

FIG. 1

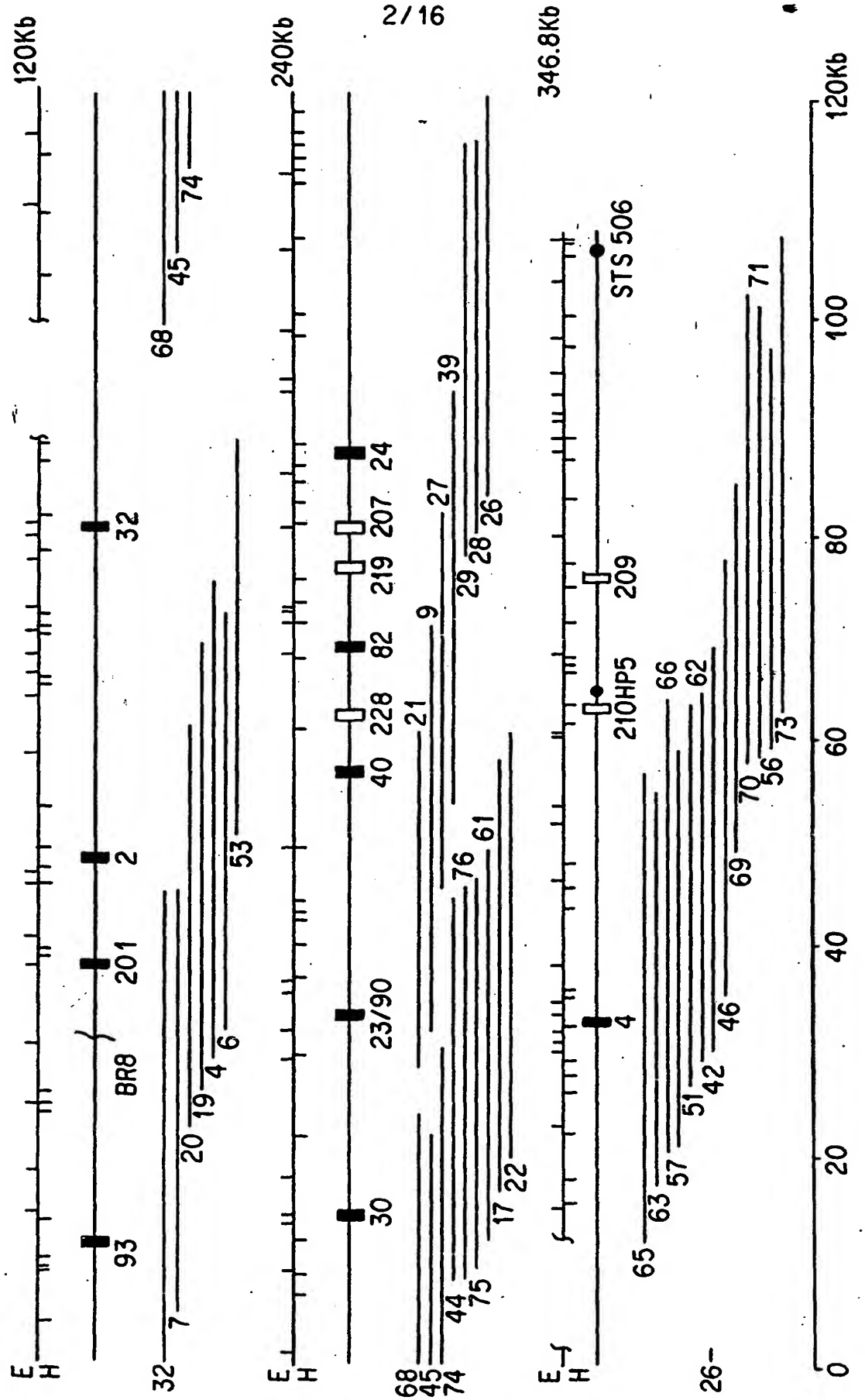

FIG. 2

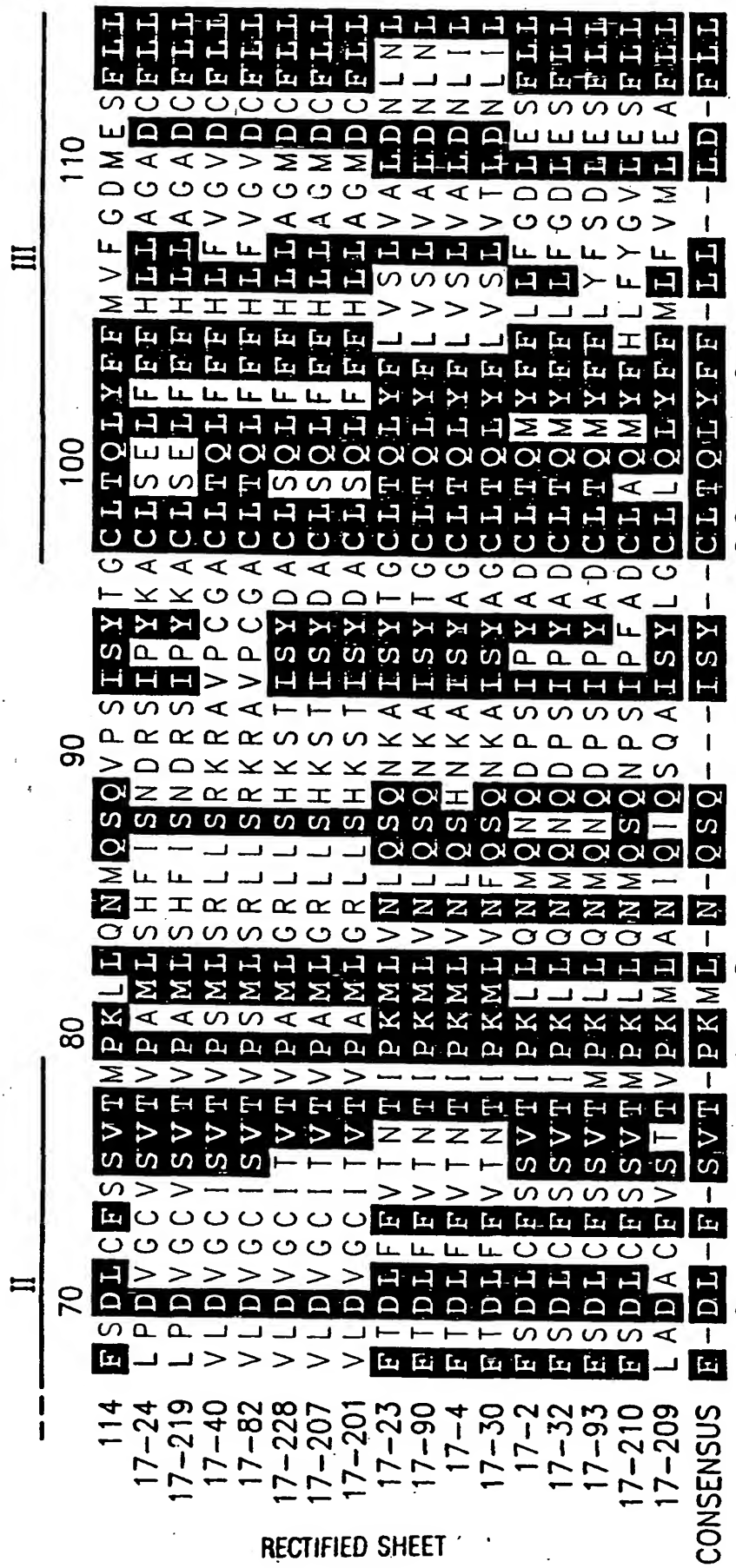

FIG. 3a

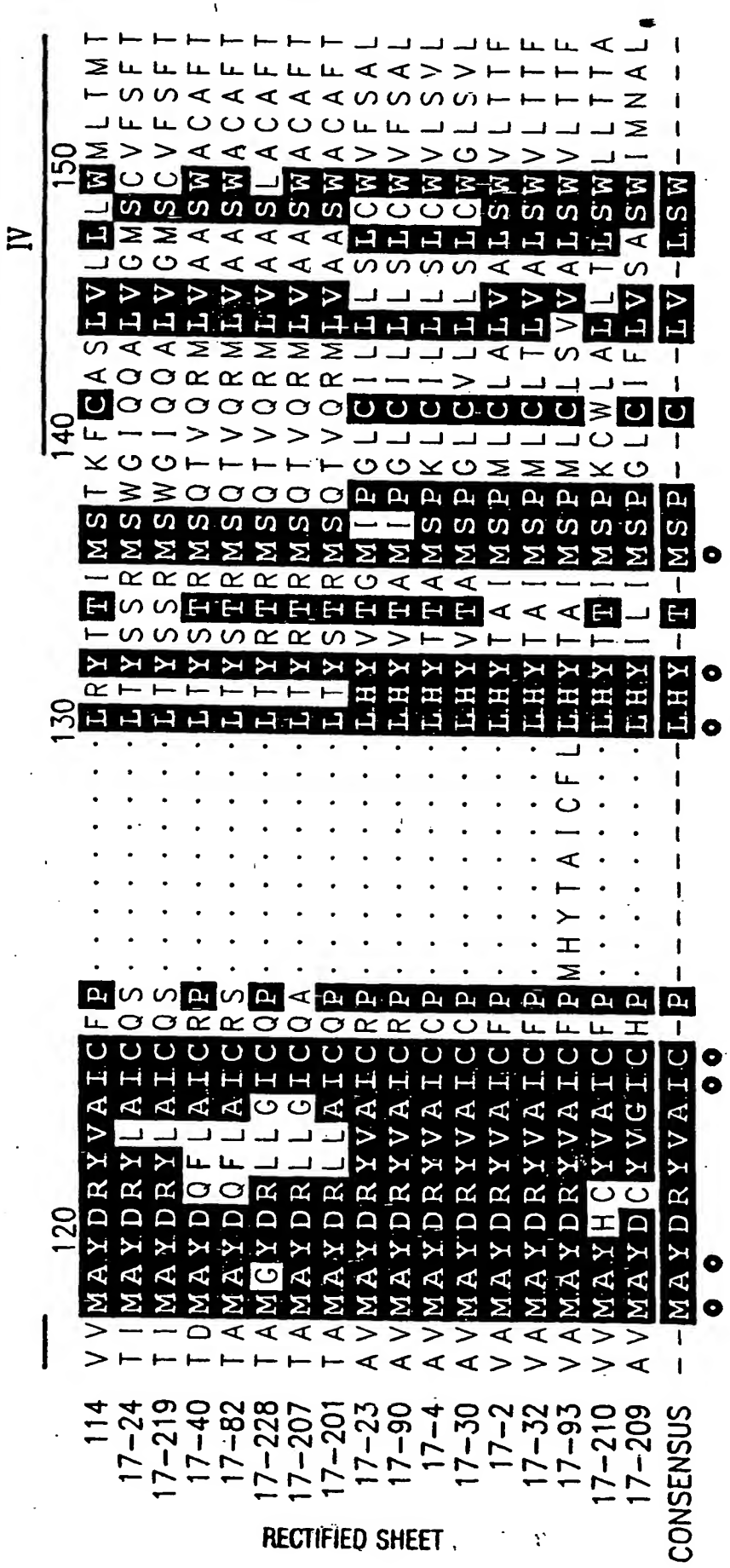

FIG. 3b

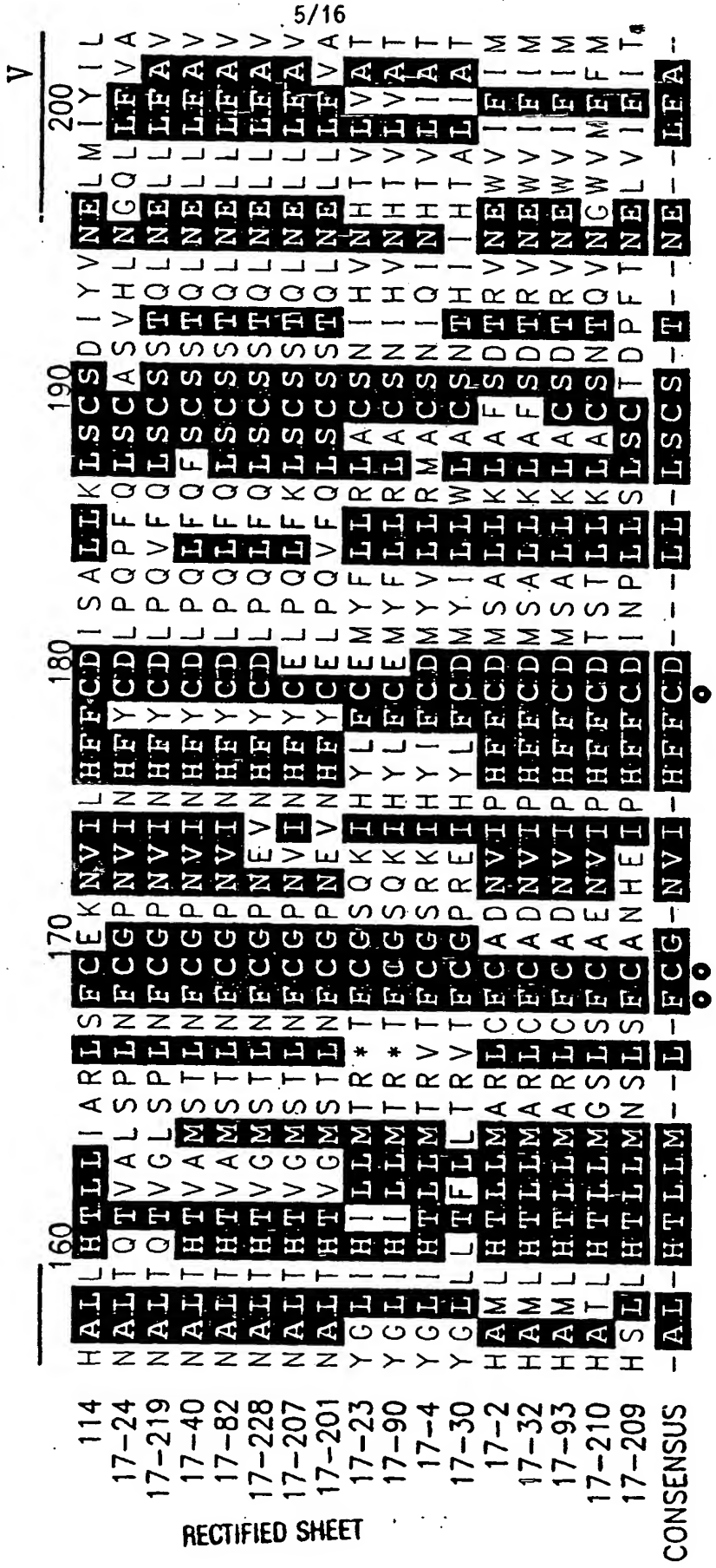

FIG. 3c

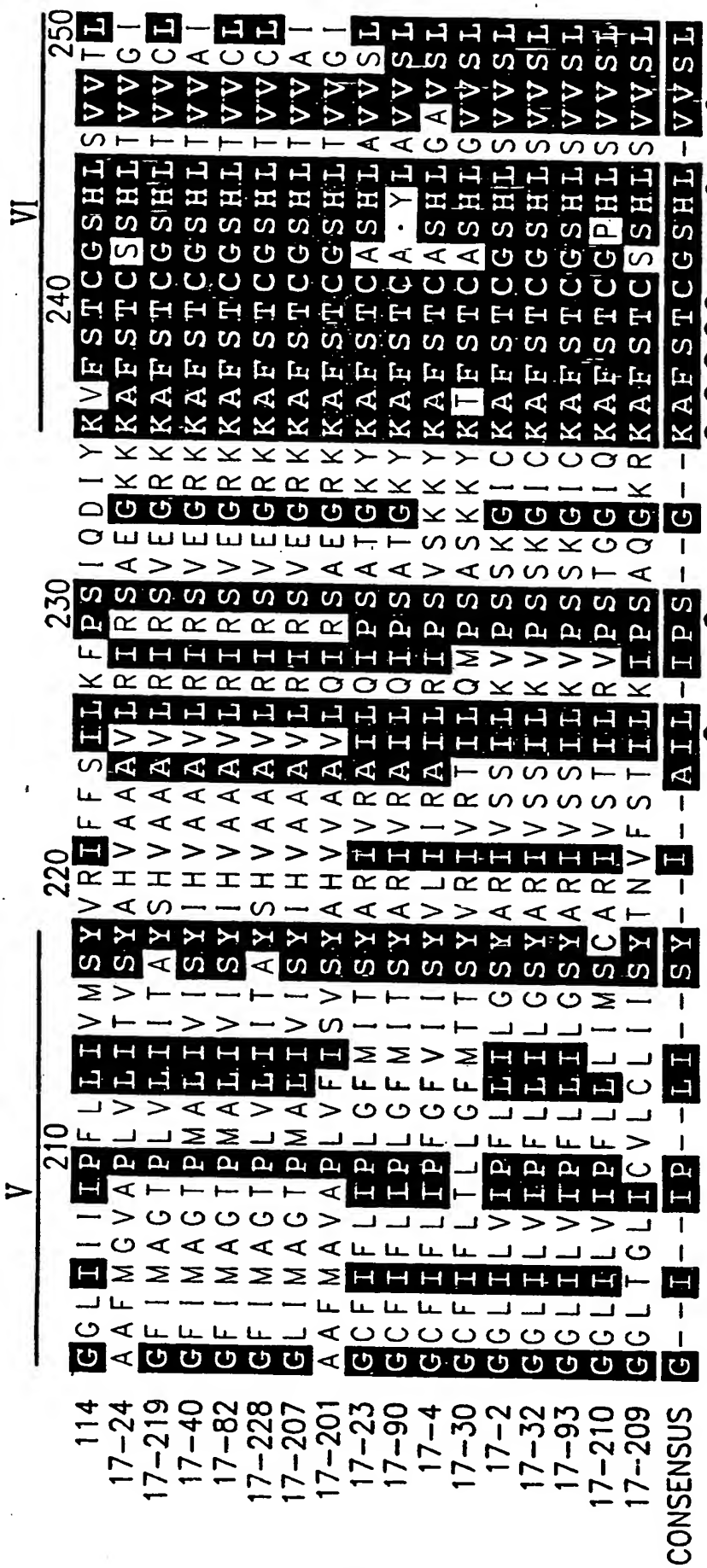
V
VI
210
220
230
240
250
114
17-24
17-219
17-40
17-82
17-228
17-207
17-201
17-23
17-90
17-4
17-30
17-2
17-32
17-93
17-210
17-209
CONSENSUS

FIG. 3d

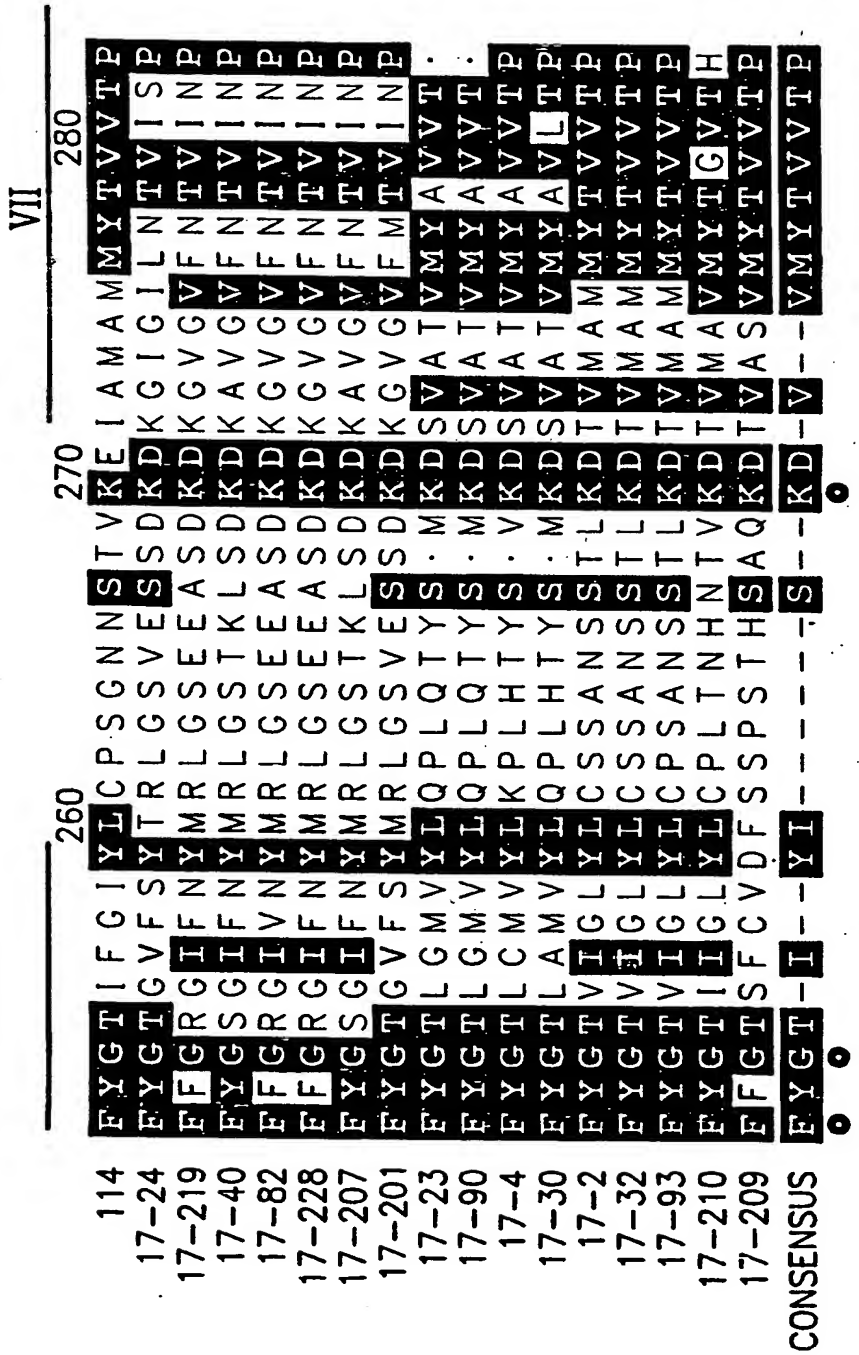
VII
260
270
280
114
17-24
17-219
17-40
17-82
17-228
17-207
17-201
17-23
17-90
17-4
17-30
17-2
17-32
17-93
17-210
17-209
CONSENSUS

FIG. 3e

17-24

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 CTGCCAGATG TGGGGTGTGT CAGTGTCACT GTCCCTGCCA  TGCTGAGCCA TTTCATATCC AACGACAGAA GCATTCCCTA 80
 81 TAAGGCCTGC CTCTCCGAGC TCTTCTTCTT CCACCTCCTG  GCTGGGGCAG ACTGCTTCTT GCTGACCATC ATGGCCTATG 160
161 ACCGCTATCT GGCCATCTGC CAGTCCCTCA CCTACAGCAG  CCGCATGAGC TGGGGAATCC AGCAAGCCCT GGTGGGCATG 240
241 TCATGTGTCT TTTCCTTCAC CAATGCACTG ACCCAAACTG  TTGCCCTGTC TCCTCTTAAC TTCTGTGGCC CCAATGTGAT 320
321 CAATCACTTC TACTGTGACC TCCCACAGCC CTTCCAGCTC  TCCTGCGCCA GTGTTCATCT CAATGGGCAG TTGCTGTTTG 400
401 TAGCAGCAGC TTTCATGGGT GTGGCCCCCT TGGTCCTCAT  CACTGTGTCC TATGCCCATG TGGCAGCTGC AGTCCTGCGA 480
481 ATCCGCTCTG CAGAGGGCAA AAAGAAAGCC TTCTCCACGT  GTAGTTCCCA CCTCACTGTG GTGGGCATCT TCTATGGGAC 560
561 GGGCGTCTTC AGCTACACAA GGCTGGGTTC AGTGGAGTCT  TCGGACAAGG ACAAGGGCAT TGGCATCCTC AACACTGTCA 640
641 TCAGCCCC                                                                              648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4a

17-219

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 CTGCCAGATG TGGGGTGTGT CAGTGTCACT GTCCCTGCCA  TGCTGAGCCA TTTCATATCC AACGACAGAA GCATTCCCTA 80
 81 TAAGGCCTGC CTCTCCGAGC TCTTCTTCTT CCACCTCCTG  GCTGGGGCAG ACTGCTTCTT GCTGACCATC ATGGCCTATG 160
161 ACCGCTATCT GGCCATCTGC CAGTCCCTCA CCTACAGCAG  CCGCATGAGC TGGGGAATCC AGCAAGCCCT GGTGGGCATG 240
241 TCATGTGTCT TTTCCTTCAC CAATGCACTG ACCCAAACTG  TTGGCCTGTC TCCTCTTAAC TTCTGTGGGC CCAATGTGAT 320
321 CAATCACTTC TACTGTGACC TCCCACAGGT CTTCCAGCTC  TCCTGCTCCA GCACCCAACT CAATGAGCTG CTGCTCTTTG 400
401 CTGTGGGTTT CATCATGGCA GGCACACCTT TGGTTCTCAT  CATCACTGCC TACAGCCACG TGGCAGCTGC AGTTCTACGA 480
481 ATCCGTTCAG TGGAGGGCCG AAAGAAGGCC TTCTCCACGT  GTGGCTCCCA CCTCACCGTG GTTTGTCTTT TCTTTGGAAG 560
561 AGGTATCTTC AACTACATGA GACTGGGTTC AGAGGAGGCT  TCAGACAAGG ATAAAGGGGT TGGAGTTTTC AACACTGTTA 640
641 TCAACCCT                                                                              648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4b

17-40

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 GTGCTGGATG TTGGGTGCAT CAGCGTCACT GTTCCATCAA  TGTTGAGTCG TCTCCTGTCC CGCAAGCGTG CAGTTCCCTG 80
 81 TGGGGCCTGC CTTACCCAGC TCTTCTTCTT CCATCTGTTC GTTGGAGTGG ACTGCTTCCT GCTGACCGAC ATGGCCTATG 160
161 ACCAATTCCT GGCCATCTGC CGGCCCCTCA CCTACAGCAC CCGCATGAGT CAGACAGTCC AGAGGATGTT GGTGGCTGCG 240
241 TCCTGGGCTT GTGCTTTCAC CAACGCACTG ACCCACACTG TGGCCATGTC CACGCTCAAC TTCTGTGGCC CCAATGTGAT 320
321 CAATCACTTC TACTGTGACC TCCCACAGCT CTTCCAGTTC TCCTGCTCCA GCACCCAACT CAATGAGCTG CTGCTTTTTG 400
401 CTGTGGGTTT TATAATGGCA GGTACCCCCA TGGCTCTCAT TGTCATCTCC TATATCCACG TGGCAGCTGC AGTCCTGCGA 480
481 ATCCGCTCTG TAGAGGGCAG GAAGAAAGCC TTCTCCACAT GTGGCTCCCA CCTCACTGTG GTTGCCATAT TCTATGGTTC 560
561 AGGTATCTTT AACTATATGC GACTGGGTTC AACCAAGCTT TCAGACAAGG ATAAAGCTGT TGGAATTTTC AACACTGTCA 640
641 TCAATCCC                                                                          648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4c

17-82

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 GTGCTGGATG TTGGGTGCAT CAGCGTCACT GTTCCATCAA  TGTTGAGTCG TCTCCTGTCC CGCAAGCGTG CAGTTCCCTG 80
 81 TGGGGCCTGC CTTACCCAGC TCTTCTTCTT CCATCTGTTC GTTGGAGTGG ACTGCTTCCT GCTGACCGCC ATGGCCTATG 160
161 ACCAATTCCT GGCCATCTGC CGGTCCCTCA CCTACAGCAC CCGCATGAGT CAGACAGTCC AGAGGATGTT GGTGGCTGCG 240
241 TCCTGGGCTT GTGCTTTCAC CAACGCACTG ACCCACACTG TGGCCATGTC CACGCTCAAC TTCTGTGGCC CCAATGTGAT 320
321 CAATCACTTC TACTGTGACC TCCCACAGCT CTTCCAGCTC TCCTGCTCCA GCACCCAACT CAATGAGCTG CTGCTTTTTG 400
401 CTGTGGGTTT TATAATGGCA GGTACCCCCA TGGCTCTCAT TGTCATCTCC TATATCCACG TGGCAGCTGC AGTCCTGCGA 480
481 ATTCGCTCTG TAGAGGGCAG GAAGAAAGCC TTCTCCACAT GTGGCTCCCA CCTCACTGTG GTTTGTCTTT TCTTTGGAAG 560
561 AGGTATCGTC AACTACATGA GACTGGGTTC AGAGGAGGCT TCAGACAAGG ATAAAGGGGT TGGAGTTTTC AACACTGTTA 640
641 TCAACCCT                                                                          648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4d

17-228

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 GTGCTGGATG TCGGATGTAT CACTGTCACT GTTCCTGCAA  TGTTGGGTCG TCTCTTGTCC CACAAGTCCA CAATTTCCTA 80
 81 TGACGCCTGC CTCTCCCAGC TCTTCTTCTT CCACCTTCTG  GCTGGGATGG ACTGCTTCCT GCTGACCGCC ATGGGCTATG 160
161 ACCGACTCCT GGGCATCTGC CAGCCCCTCA CCTACAGGAC  CCGCATGAGT CAGACAGTCC AGAGGATGTT GGTGGCTGCG 240
241 TCCTTGGCTT GTGCCTTCAC CAACGCACTG ACCCACACTG  TGGGCATGTC CACGCTCAAC TTCTGTGGGC CCAATGAGGT 320
321 CAATCACTTC TACTGTGACC TCCCACAGCT CTTCCAGCTC  TCCTGCTCCA GCACCCAACT CAATGAGCTG CTGCTCTTTG 400
401 CTGTGGGTTT CATCATGGCA GGCACACCTT TGGTTCTCAT  CATCACTGCC TACAGCCACG TGGCAGCTGC AGTTCTACGA 480
481 ATCCGTTCAG TGGAGGGCCG AAAGAAGGCC TTCTCCACGT  GTGGCTCCCA CCTCACCGTG GTTTGTCTTT TCTTTGGAAG 560
561 AGGTATCTTC AACTACATGA GACTGGGTTC AGAGGAGGCT  TCAGACAAGG ATAAAGGGGT TGGAGTTTTC AACACTGTTA 640
641 TCAACCCT                                                                          648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4e

17-207

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 GTGCTGGATG TCGGATGTAT CACTGTCACT GTTCCTGCAA  TGTTGGGTCG TCTCTTGTCC CACAAGTCCA CAATTTCCTA 80
 81 TGACGCCTGC CTCTCCCAGC TCTTCTTCTT CCACCTTCTG  GCTGGGATGG ACTGCTTCCT GCTGACCGCC ATGGCCTATG 160
161 ACCGACTCCT GGGCATCTGT CAGGCCCTCA CCTACAGGAC  CCGCATGAGT CAGACAGTCC AGAGGATGTT GGTGGCTGCG 240
241 TCCTGGGCTT GTGCTTTCAC CAACGCACTG ACCCACACTG  TGGGCATGTC CACGCTCAAC TTCTGTGGGC CCAATGTGAT 320
321 CAATCACTTC TACTGTGACC TCCCACAGCT CTTCAAGCTC  TCCTGCTCCA GCACCCAACT CAATGAGCTG CTGCTTTTTG 400
401 CTGTGGGTTT AATAATGGCA GGTACCCCCA TGGCTCTCAT  TGTCATCTCC TATATCCACG TGGCAGCTGC AGTCCTGCGA 480
481 ATTCGCTCTG TAGAGGGCAG GAAGAAAGCC TTCTCCACAT  GTGGCTCCCA CCTCACTGTG GTTGCCATAT TCTATGGTTC 560
561 AGGTATCTTT AACTATATGC GACTGGGTTC AACCAAGCTT  TCAGACAAGG ATAAAGCTGT TGGAATTTTC AACACTGTCA 640
641 TCAATCCC                                                                          648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4f

17-201

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
    1 GTGCTGGATG TCGGATGTAT CACTGTCACT GTTCCTGCAA  TGTTGGGTCG TCTCTTGTCC CACAAGTCCA CAATTTCCTA 80
   81 TGACGCCTGC CTCTCCCAGC TCTTCTTCTT CCACCTTCTG  GCTGGGATGG ACTGCTTCCT GCTGACCGCC ATGGCCTATG 160
  161 ACCGACTCCT GGcCATCTGC CAGcCCCTCA CCTACAGcAC  CCGCATGAGT CAGACAGTCC AGAGGATGTT GGTGGCTGCG 240
  241 TCCTGGGCTT GTGCCTTCAC CAACGCACTG ACCCACACTG  TGGGCATGTC CACGCTCAAC TTCTGTGGGC CCAATGAGGT 320
  321 CAATCACTTC TACTGTGAAC TCCCACAGGT CTTCCAGCTC  TCCTGCTCCA GCACCCAACT CAATGAGCTG CTGCTCTTTG 400
  401 TAGCAGCAGC CTTCATGGCT GTGGCACCCT TGGTCTTCAT  CAGTGTGTCC TATGCCCATG TGGTAGCTGC TGTGCTGCAA 480
  481 ATCCGCTCTG CTGAGGGCAG AAAGAAGGCC TTCTCCACAT  GTGGCTCCCA CCTCACTGTG GTGGGCATCT TCTATGGGAC 560
  561 AGGTGTCTTC AGCTACATGA GGCTGGGTTC AGTGGAATCT  TCAGACAAGG ATAAGGGGGT TGGGGTTTTC ATGACTGTGA 640
  641 TCAACCCC                                                                              648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4g

17-23

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
    1 TTCACTGACC TCTTCTTTGT CACCAACACA ATCCCCAAGA  TGCTGGTGAA CCTCCAGTCC CAGAACAAAG CCATCTCCTA 80
   81 CACAGGGTGT CTGACACAGC TCTACTTCCT GGTCTCCTTG  GTGGCCCTGG ACAACCTCAA CCTGGCCGTG ATGGCGTATG 160
  161 ATCGCTATGT GGCCATCTGC CGTCCCCTCC ACTATGTCAC  AGGCATGATC CCTGGGCTCT GTATCTTGCT CCTCTCCTTG 240
  241 TGTTGGGTGT TCTCTGCCCT CTATGGCCTC ATCCATATCC  TCCTCATGAC CAGGgTGACC TTCTGTGGGT CTCAAAAGAT 320
  321 CCACTACCTC TTCTGTGAGA TGTACTTCCT GCTAAGGCTG  GCATGTTCCA ACATCCACGT CAACCACACA GTACTGGTTG 400
  401 CCACGGGCTG CTTCATCTTC CTCATCCCCT TAGGTTTCAT  GATCACATCC TACGCCCGCA TTGTCAGAGC CATCCTCCAA 480
  481 ATACCCTCAG CCACTGGGAA GTACAAAGCC TTCTCCACCT  GTGCTTCCCA TTTGGCTGTG GTCTCCCTCT TCTATGGGAC 560
  561 TCTGGGTATG GTGTACCTGC AGCCCCTCCA AACCTACTCC  ATGAAGGACT CAGTAGCCAC AGTGATGTAT GCGGTGGTGA 640
  641 CGCC                                                                                  644
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4h

17-90

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 TTCACTGACC TCTTCTTTGT CACCAACACA ATCCCCAAGA  TGCTGGTGAA CCTCCAGTCC CAGAACAAAG CCATCTCCTA 80
 81 CACAGGGTGT CTGACACAGC TCTACTTCCT GGTCTCCTTG  GTGGCCCTGG ACAACCTCAA CCTGGCCGTG ATGGCGTATG 160
161 ATCGCTATGT GGCCATCTGC CGTCCCCTCC ACTATGTCAC  AGCCATGATC CCTGGGCTCT GTATCTTGCT CCTCTCCTTG 240
241 TGTTGGGTGT TCTCTGCCCT CTATGGCCTC ATCCATATCC  TCCTCATGAC CAGGgTGACC TTCTGTGGGT CTCAAAAGAT 320
321 CCACTACCTC TTCTGTGAGA TGTACTTCCT GCTAAGGCTG  GCATGTTCCA ACATCCACGT CAACCACACA GTACTGGTTG 400
401 CCACGGGCTG CTTCATCTTC CTCATCCCCT TAGGTTTCAT  GATCACATCC TACGCCCGCA TTGTCAGAGC CATCCTCCAA 480
481 ATACCCTCAG CCACTGGGAA GTACAAAGCC TTCTCCACCT  GTGCTTATTT GGCTGTGGTC TCCCTCTTCT ATGGGACTCT 560
561 GGGTATGGTG TACCTGCAGC CCCTCCAAAC CTACTCCATG  AAGGACTCAG TAGCCACAGT GATGTATGCG GTGGTGACGC 640
641 C                                                                                      641
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4i

17-4

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 TTCACTGACC TCTTCTTTGT CACCAACACA ATCCCCAAGA  TGCTGGTGAA CCTCCAGTCC CATAACAAAG CCATCTCCTA 80
 81 TGCAGGGTGT CTGACACAGC TCTACTTCCT GGTCTCCTTG  GTGGCCCTGG ACAACCTCAT CCTGGCTGTG ATGGCATATG 160
161 ACCGCTATGT GGCCATCTGC TGCCCCCTCC ACTACACCAC  AGCCATGAGC CCTAAGCTCT GTATCTTACT CCTTTCCTTG 240
241 TGTTGGGTCC TATCCGTCCT CTATGGCCTC ATACACACCC  TCCTCATGAC CAGAGTGACC TTCTGTGGGT CACGAAAAAT 320
321 CCACTACATC TTCTGTGAGA TGTATGTATT GCTGAGGATG  GCATGTTCCA ACATTCAGAT TAATCACACA GTGCTGATTG 400
401 CCACAGGCTG CTTCATCTTC CTCATTCCCT TTGGATTCGT  GATCATTTCC TATGTGCTGA TTATCAGAGC CATCCTCAGA 480
481 ATACCCTCAG TCTCTAAGAA ATACAAAGCC TTCTCCACCT  GTGCCTCCCA TTTGGGTGCA GTCTCCCTCT TCTATGGGAC 560
561 ACTTTGTATG GTATACCTAA AGCCCCTCCA TACCTACTCT  GTGAAGGACT CAGTAGCCAC AGTGATGTAT GCTGTGGTGA 640
641 CACCC                                                                                  645
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4j

17-30

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 TTCACTGACC TCTTCTTTGT CACCAACACA ATCCCCAAGA  TGCTGGTGAA CTTCCAGTCC CAGAACAAAG CCATCTCCTA 80
 81 TGCAGGGTGT CTGACACAGC TCTACTTCCT GGTCTCCTTG  GTGACCCTGG ACAACCTCAT CCTGGCCGTG ATGGCGTATG 160
161 ATCGCTATGT GGCCATCTGC TGCCCCCTCC ACTATGTCAC  AGCCATGAGC CCTGGGCTCT GTGTCTTGCT CCTCTCCTTG 240
241 TGTTGGGGGC TGTCTGTTCT CTATGGCCTC CTCCTCACCT  TCCTCCTGAC CAGGGTGACC TTCTGTGGGC CTCGAGAGAT 320
321 CCACTACCTC TTCTGTGACA TGTACATCCT GCTGTGGCTG  GCATGTTCCA ACACCCACAT CATTCACACA GCATTGATTG 400
401 CCACTGGCTG CTTCATCTTC CTCACCCTCT TAGGGTTCAT  GACCACATCC TATGTACGTA TTGTCAGAAC CATCCTTCAA 480
481 ATGCCCTCGG CCTCTAAGAA ATACAAAACT TTCTCTACCT  GTGCCTCGCA TTTGGGTGTG GTCTCCCTCT TTTATGGGAC 560
561 GCTTGCTATG GTGTACCTGC AGCCCCTCCA TACCTACTCC  ATGAAGGACT CAGTAGCCAC AGTGATGTAT GCTGTGCTGA 640
641 CACCT                                                                              645
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4k

17-2

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 TTCTCTGACC TCTGCTTCTC TTCCGTGACC ATTCCCAAGT  TGTTACAGAA CATGCAGAAC CAGGACCCAT CCATCCCCTA 80
 81 TGCGGACTGC CTGACCCAAA TGTACTTCTT CCTGTTATTT  GGAGACCTGG AGAGCTTCCT CCTTGTGGCC ATGGCCTATG 160
161 ACCGCTATGT GGCCATCTGC TTCCCCCTGC ACTACACCGC  CATCATGAGC CCCATGCTCT GTCTCGCCCT GGTGGCGCTG 240
241 TCCTGGGTGC TGACCACCTT CCATGCCATG TTACACACTT  TACTCATGGC CAGGTTGTGT TTTTGTGCAG ACAATGTGAT 320
321 CCCCCACTTT TTCTGTGATA TGTCTGCTCT GCTGAAGCTG  GCCTTCTCTG ACACTCGAGT TAATGAATGG GTGATATTTA 400
401 TCATGGGAGG GCTCATTCTT GTCATCCCAT TCCTACTCAT  CCTTGGGTCC TATGCAAGAA TTGTCTCCTC CATCCTCAAG 480
481 GTCCCTTCTT CTAAGGGTAT CTGCAAGGCC TTCTCTACTT  GTGGCTCCCA CCTGTCTGTG GTGTCACTGT TCTATGGAAC 560
561 CGTTATTGGT CTCTACTTAT GCTCATCAGC TAATAGTTCT  ACTCTAAAGG ACACTGTCAT GGCTATGATG TACACTGTGG 640
641 TGACCCCC                                                                           648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4l

17-32

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 TTCTCTGACC TCTGCTTCTC TTCCGTGACC ATTCCCAAGT  TGTTACAGAA CATGCAGAAC CAGGACCCAT CCATCCCCTA 80
 81 TGCGGACTGC CTGACCCAAA TGTACTTCTT CCTGTTATTT  GGAGACCTGG AGAGCTTCCT CCTTGTGGCC ATGGCCTATG 160
161 ACCGCTATGT GGCCATCTGC TTCCCCCTGC ACTACACCGC  CATCATGAGC CCCATGCTCT GTCTCACCCT GGTGGCGCTG 240
241 TCCTGGGTGC TGACCACCTT CCATGCCATG TTACACACTT  TACTCATGGC CAGGTTGTGT TTTTGTGCAG ACAATGTGAT 320
321 CCCCCACTTT TTCTGTGATA TGTCTGCTCT GCTGAAGCTG  GCCTTCTCTG ACACTCGAGT TAATGAATGG GTGATATTTA 400
401 TCATGGGAGG GCTCATTCTT GTCATCCCAT TCCTACTCAT  CCTTGGGTCC TATGCAAGAA TTGTCTCCTC CATCCTCAAG 480
481 GTCCCTTCTT CTAAGGGTAT CTGCAAGGCC TTCTCTACTT  GTGGCTCCCA CCTGTCTGTG GTGTCACTGT TCTATGGAAC 560
561 CGTTATTGGT CTCTACTTAT GCTCATCAGC TAATAGTTCT  ACTCTAAAGG ACACTGTCAT GGCTATGATG TACACTGTGG 640
641 TGACCCCC                                                                              648
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4m

17-93

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 TTCTCTGACC TCTGCTTTTC CTCAGTCACA ATGCCCAAAT  TGCTGCAGAA CATGCAGAAC CAAGACCCAT CCATCCCCTA 80
 81 TGCAGACTGT CTGACCCAAA TGTACTTCTT CTTGTATTTT  TCGGATCTAG AGAGCTTCCT CCTTGTGGCC ATGGCCTATG 160
161 ACCGCTATGT GGCCATCTGC TTCCCCATGC ACTACACCGC  CATCTGCTTC CTCCTGCACT ACACCGCCAT CATGAGCCCC 240
241 ATGCTCTGTC TCTCCGTGGT GGCGCTGTCC TGGGTGCTGA  CCACCTTCCA TGCCATGTTA CACACTTTAC TCATGGCCAG 320
321 GTTGTGTTTT TGTGCAGACA ATGTGATCCC CCACTTTTTC  TGTGATATGT CTGCTCTGCT GAAGCTGGCC TGCTCTGACA 400
401 CTCGAGTTAA TGAATGGGTG ATATTTATCA TGGGAGGGCT  CATTCTTGTC ATCCCATTCC TACTCATCCT TGGGTCCTAT 480
481 GCAAGAATTG TCTCCTCCAT CCTCAAGGTC CCTTCTTCTA  AGGGTATCTG CAAGGCCTTC TCTACTTGTG GCTCCCACCT 560
561 CTCTGTGGTG TCACTGTTCT ATGGGACCGT TATTGGTCTC  TACTTATGCC CATCAGCTAA TAGTTCTACT CTAAAGGACA 640
641 CTGTCATGGC TATGATGTAC ACTGTGGTGA CCCCT                                                 675
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4n

17-210

```
        |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 TTCTCTGACC TCTGCTTTTC CTCGGTCACA ATGCCCAAAT TGCTGCAGAA CATGCAGAGC CAAAACCCAT CCATCCCCTT 80
 81 TGCGGACTGC CTGGCTCAGA TGTACTTTCA TCTGTTTTAT GGAGTTCTGG AGAGCTTCCT CCTTGTGGTC ATGGCTTATC 160
161 ACTGCTATGT GGCTATTTGC TTTCCTCTGC ACTACACCAC TATCATGAGC CCCAAGTGTT GGCTTGCTCT GCTGACACTC 240
241 TCCTGGCTGT TGACCACTGC CCATGCCACG TTGCACACCT TGCTTATGGG CAGCCTGTCC TTTTGTGCTG AGAATGTGAT 320
321 TCCTCACTTT TTCTGTGATA CATCTACCTT GTTGAAGCTG GCCTGCTCCA ACACGCAAGT CAAtGGGTGG GTGATGTTTT 400
401 TCATGGGCGG GCTCATCCTT GTCATCCCAT TCCTACTCCT CATCATGTCC TGTGCAAGAA TCGTCTCCAC CATCCTCAGG 480
481 GTCCCTTCCA CTGGGGGCAT CCAGAAGGCT TTCTCCACCT GTGGCCCCCA CCTCTCTGTG GTGTCTCTCT TCTATGGGAC 560
561 AATTATTGGT CTCTACTTGT GCCCATTGAC GAATCATAAC ACTGTGAAGG ACACTGTCAT GGCTGTGATG TACACTGGGG 640
641 TGACCCAC                                                                            648
        |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4o

17-209

```
        |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
  1 CTTGCAGATG CCTGCTTTGT GTCCACCACA GTCCCTAAGA TGCTGGCAAA CATACAGATC CAGAGTCAGG CCATCTCCTA 80
 81 CTTAGGGTGT CTACTACAGT TGTATTTTTT CATGTTATTT GTGATGCTGG AGGCATTCCT CTTGGCGGTC ATGGCCTATG 160
161 ACTGCTACGT GGGCATATGC CACCCACTTC ATTACATTCT GATCATGAGC CCTGGGCTCT GCATCTTCCT CGTGTCTGCA 240
241 TCCTGGATCA TGAATGCCCT CCACTCCCTT CTACACACAC TTCTGATGAA CAGCCTGTCC TTCTGCGCAA ACCATGAGAT 320
321 CCCACACTTC TTCTGTGACA TCAATCCCCT CCTGAGTCTG TCCTGCAcAG ACCCCTTCAC CAATGAGCTG GTGATCTTCA 400
401 TCACTGGGGG TCTCACAGGA CTCATTTGTG TGCTTTGCCT GATTATCTCT TACACGAACG TTTTCTCGAC CATCCTGAAG 480
481 ATCCCATCAG CTCAGGGGAA GCGGAAAGCC TTTTCCACCT GCAGCTCTCA TCTCTCCGTG GTCTCTCTCT TCTTTGGGAC 560
561 TTCTTTTTGT GTTGATTTCA GTTCTCCCTC AACCCACTCG GCCCAGAAGG ACACAGTTGC ATCAGTGATG TACACAGTGG 640
641 TAACTCCA                                                                            648
        |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

FIG. 4p

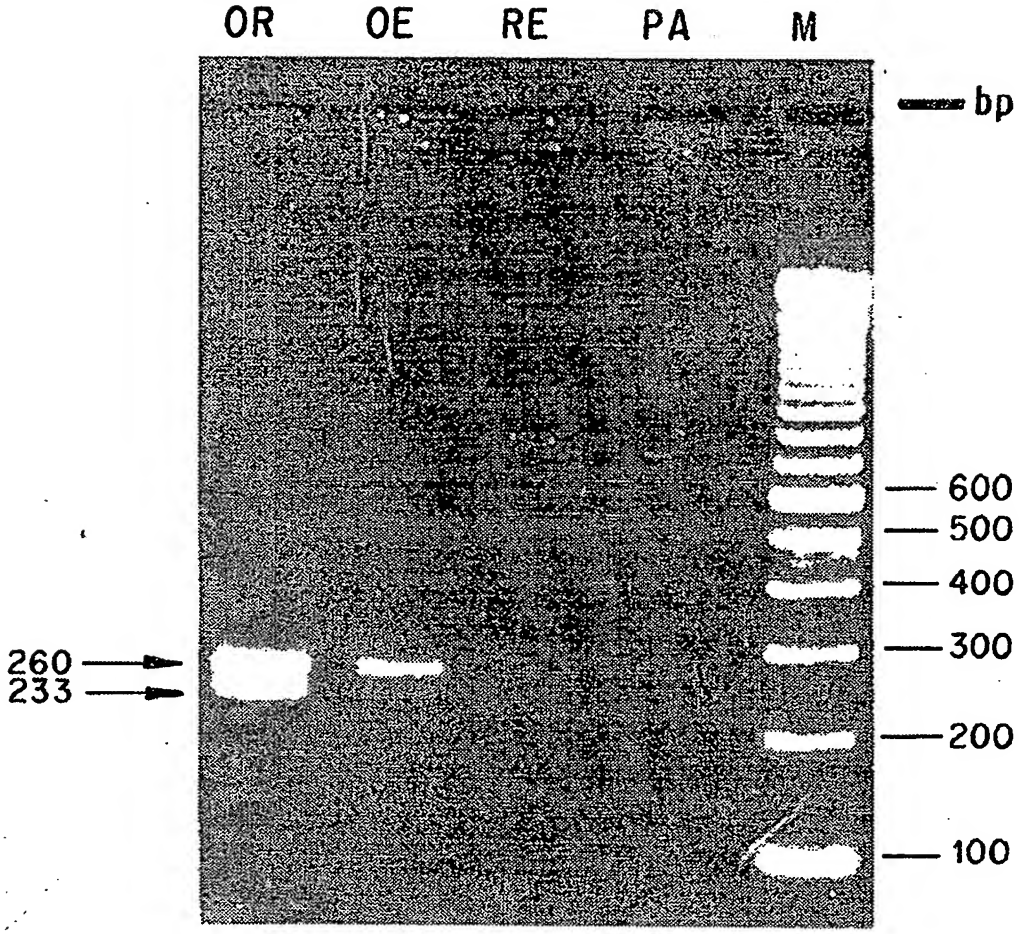

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/14554

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 70.1, 70.3, 71.1, 240.1, 240.2, 320.1; 530/300, 350; 536/23.1, 23.5, 24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Nature, Volume 355, issued 30 January 1992, M. Parmentier et al, "Expression of Members of the Putative Olfactory Receptor Gene Family in Mammalian Germ Cells", pages 453-45, see entire documents, especially page 454.	1-8
Y	Nature, Volume 361, issued 28 January 1993, K. Raming et al, "Cloning and Expression of Odorant Receptors", pages 353-356, see entire document, especially page 354.	1-8
Y	Cell, Volume 65, issued 05 April 1991, L. Buck et al, "A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition", pages 175-187, see entire document, especially pages 179-180.	1-8

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
19 MARCH 1995

Date of mailing of the international search report
03 APR 1995

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
ARIE M. MICHELSOHN, PH.D.
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/14554

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim N .
A	Trends in Neuroscience, Volume 14, No. 7, issued 1991, S. Firestein, "A Noseful of Odor Receptors", pages 270-272.	1-8
Y	Cytogenet. Cell Genet., Volume 63, issued 1993, S. Schurmans et al, "The OLFR1 Gene Encoding the HGMP07E Putative Olfactory Receptor Maps to the 17p13-->p12 Region of the Human Genome and Reveals and *Msp*I Restriction Fragment Length Polymorphism", pages 200-204, see entire document.	1-8
Y	WO, A, 92/17585 (AXEL ET AL) 15 October 1992, see entire document.	1-8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/14554

B x I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please see Extra Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-8

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ N protest accompanied the payment f additi nal search fees.

INTERNATIONAL SEARCH REPORT

In...national application No.
PCT/US94/14554

BOX II OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING:

I. Claims 1-8, directed to nucleotide sequences and corresponding protein sequences, host cells and recombinant method of producing the corresponding protein sequences, classified in U.S. Class 435, subclass 69.1 and Class 530, subclass 350, and others.

II. Claim 9, directed to an antibody, classified in U.S. Class 530, subclass 387.9.

Groups I and II are materially distinct compositions of matter that are distinguished, each from the other, by their special technical features: the polynucleotides and corresponding polypeptides, host cells and method of producing the corresponding proteins of Group I, and the antibody of Group II, have materially different structures and functions. Accordingly, the claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C07H 21/00, 21/04; C07K 2/00, 14/00; C12N 5/10, 5/16, 7/01, 15/00, 15/10, 15/11, 15/12, 15/63; C12P 1/00, 21/00, 21/02

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/69.1, 70.1, 70.3, 71.1, 240.1, 240.2, 320.1; 530/300, 350; 536/23.1, 23.5, 24.31

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS; STN files Biosis, Medline, EMBASE, CA, WPIDS; sequence search through Intelligenetics; databanks EMBL-NEW, GenBank 86 & NEW, UEMBL. Search terms for STN search: (olfactory(w)receptor#) and clon?, cDNA, DNA, amino acid sequence#, genomic DNA, recombinant DNA, nucleic acid#.